Adaptive Hypothesis Testing using Wavelets

Spokoiny, V.G.

Weierstrass Institute for Applied Analysis and Stochastics,
Mohrenstr. 39, 10117 Berlin
and
Institute for Information Transmission Problems,
Ermolovoy 19, Moscow 101447, Russia

1980 *Mathematics Subject Classification* (1985 *Revision*). Primary 62G10; Secondary 62G20.

Key words and phrases. signal detection, minimax hypothesis testing, nonparametric alternative, error probabilities, wavelet decomposition, thresholding, adaptive test .

The author thanks O. Lepski, D.Picard, G.Kerkyachrian, A. Juditsky, A. Nemirovskii and M. Neumann for the helpful remarks and discussion..

ABSTRACT. The present paper continues studying the problem of minimax nonparametric hypothesis testing started in Lepski and Spokoiny (1995).

The null hypothesis assumes that the function observed with a noise is identically zero i.e. no signal is present. The alternative is composite and minimax: the function is assumed to be separated away from zero in an integral (L_2 –) norm and also to possess some smoothness properties.

The minimax rate of testing for this problem was evaluated by Ingster for the case of Sobolev smoothness classes. Then this problem was studied by Lepski and Spokoiny in the sutiation of an alternative with inhomogeneous smoothness properties that leads to considering Besov smoothness classes.

But for both cases the optimal rate and the structure of optimal (in rate) tests depends on smoothness parameters which are usually unknown in practical applications.

In this paper the problem of adaptive (assumption free) testing is considered. It is shown that the adaptation without loss of efficiency is impossible. An extra (log log)-factor is nonsignificant but unavoidable payment for the adaptation.

A simple adaptive test based on wavelet technique is constructed which is nearly minimax for a wide range of Besov classes.

1. Introduction

The present paper continues the study started in Lepski and Spokoiny (1995) on minimax nonparametric hypothesis testing. The reader is referred to that paper and to Ingster (1993) for the detailed historical background to this problem. We recall only the main points.

We suppose that we are given data

$$dX(t) = f(t)dt + \varepsilon dW(t), \qquad 0 \le t \le 1, \tag{1.1}$$

where f is an unknown function and W is a standard Wiener process.

We wish to test the null hypothesis $H_0 : f \equiv 0$ against the composite nonparametric alternative that the function f is separated away from zero in L_2-norm, $\|f\| \ge r(\varepsilon)$, and also f possesses some smoothness properties.

The problem is to describe the minimal (optimal) rate for the distance $r(\varepsilon)$ for which testing with prescribed error probabilities is still possible. The result depends heavily on which kind of smoothness assumption we impose. For the case of Sobolev-type functional classes this problem was explicitly solved by Ingster (1982, 1993) and Ermakov (1990). It turned out that the optimal rate $r(\varepsilon)$ for testing differs from the rate for estimation: if σ is the smoothness parameter, then

$$r(\varepsilon) = \varepsilon^{\frac{4\sigma}{4\sigma+1}}.$$

The case of Besov functional classes $B^{\sigma}_{p,q}$ with $p < 2$ was considered in Lepski and Spokoiny (1995). This case is not only of theoretical interest. It corresponds to the situation when functions from the alternative set are of inhomogeneous smoothness properties that is very important both for theory and practice. The optimal rate was proved to be

$$r(\varepsilon) = \varepsilon^{\frac{4\sigma''}{4\sigma''+1}}$$

where $\sigma'' = \sigma - 1/(2p) + 1/4$. The rate-optimal test constructed in that paper makes heavy use the pointwise adaptive procedure proposed in Lepski et al.(1994) and developed in Lepski and Spokoiny (1995).

But the practical applications of this test or that of proposed by Ingster meet the crucial problem: the structure of the test uses knowledge of the smoothness parameters σ, p which are typically unknown.

The present paper treats the problem of adaptive (assumption free) testing. The goal is to propose a test which does not use any information about smoothness properties of the function f but which is at least nearly optimal.

The problem of adaptive nonparametric estimation is now well developed. We mention here the paper by Efroimovich and Pinsker (1984), Poljak and Tsybakov (1990), Golubev (1987), Lepski (1990). The reader is referred to Donoho and Johnstone (1992) and Marron (1987) for further information on this problem in the considered context. Note that global adaptive estimation is possible without loss of efficiency and can be performed even in an optimal way (up to constant).

Another interesting phenomena was discovered by Lepski (1990) and then Brown and Low (1992): for some statistical estimation problem an adaptive estimation without loss of efficiency is impossible. Such a problem is, for instance, the estimation of a function f at a given point t_0. The optimal adaptive rate was also calculated in Lepski (1990) which appeared to be worse by an extra log-factor.

Then in Lepski and Spokoiny (1995) an optimal adaptive risk was calculated and optimal adaptive procedure was constructed to this problem.

In the present paper it is shown that adaptive testing also leads to some loss of efficiency but in this case with an extra (log log)-factor. The difference with the preceding case is explained mostly by the structure of the loss function (it is bounded in the hypothesis testing problem). But the related consideration seems to be more involved.

The rate-optimal adaptive test is also presented. We use for the construction the wavelet technique which provides very useful tools for studying the problem under consideration.

Finally we describe one more test which has a slightly worse performance but which of relatively simple and obvious structure. The consideration of this test is motivated by the fact that it is a direct analogue of the nonlinear thresholding wavelet procedure, see Donoho et al. (1994).

The paper is organized as follows. In the next section we state the testing problem for the model "signal + white noise". Then we translate this problem into the sequence space and restate the problem in terms of empirical wavelet coefficients.

In Section 3 we present the results on the minimax rate of testing and propose a test which is rate optimal. This test is based on wavelet decomposition and its construction involves only one external parameter. But the value of this external parameter is determined by the smoothness assumptions on the alternative set.

In the next section the problem of adaptive testing is discussed. We define the notion of adaptive rate of testing and describe the structure of adaptive tests which are rate optimal or nearly optimal.

In Section 5 we translate the results back into the nonparametric function model.

The proofs are postponed to the last sections.

2. Hypothesis testing problem

In this section we recall the definition of the minimax hypothesis testing problem. We assume the model (1.1). Our goal is to test the null hypothesis H_0 that the function f is identically zero,

$H_0 : \qquad f \equiv 0\,.$

The alternative is composite and nonparametric. First we assume that the alternative set is separated away from null in the L_2-norm,

$$\|f\| \geq r(\varepsilon) \tag{2.1}$$

where $\|f\|^2 = \int_0^1 f^2(t)dt$ and the radius $r(\varepsilon)$ qualifies sensitivity of testing.

Notice, however, that only the assumption (2.1) is not enough for testing for any $r(\varepsilon)$. This fact was stated by Ingster (1982) but it is intuitively clear: without special assumptions one is not able to distinguish between a function and a noise.

To bypass this problem, some smoothness assumptions are additionally imposed on the function f. Together with (2.1) these assumptions allow to test H_0 consistently or with prescribed error probabilities.

In the papers by Ingster, Ermakov, Suslina, it is assumed that f belongs to some ellipsoidal body in the space of coefficients for some orthonormal system in the function space.

In the present paper, similarly to the paper by Lepski and Spokoiny (1995), we treat the case of Besov balls $B^\sigma_{p,q}(M) = \{f : \|f\|_{B^\sigma_{p,q}} \leq M\}$ where

$$\|f\|_{B^\sigma_{p,q}} = \begin{cases} \|f\|_p + \left[\int_0^1 h^{-\sigma q} \|\mathrm{osc}\ f(\cdot,h)\|_p^q \frac{dh}{h}\right]^{1/q} & \text{if } q < \infty, \\ \|f\|_p + \sup_{0\leq h\leq 1} h^{-\sigma}\|\mathrm{osc}\ f(\cdot,h)\|_p & \text{if } q = +\infty. \end{cases}$$

Here $\|f\|_p$ is the L_p-norm, $\|f\|_p^p = \int_0^1 |f|^p$. The local oscillation $\mathrm{osc}\ f(x,h)$ of f is defined as

$$\mathrm{osc}\ f(x,h) = \inf \sup_{|y-x|\leq h} |f(y) - P(y)|$$

the infimum is taken over all polynomials of order r, where r is an integer greater or equal to σ and sup in x, y is restricted to the interval $[0,1]$.

This leads to the following nonparametric alternative

$H_1 : \qquad \|f\|_{B^\sigma_{p,q}} \leq M,\ \|f\| \geq r(\varepsilon)\,.$

2.1. A test and its power function. Minimax rate of testing

Let φ be a test i.e. a function of observation process $X(t)$, $t \in [0,1]$ with two values. The event $\varphi = 0$ is treated as accepting the hypothesis H_0 and $\varphi = 1$ means that the hypothesis is rejected.

We measure the quality of the test φ by its error probabilities. Given positive nimbers α_0 and β_0 with $\alpha_0 + \beta_0 < 1$, we search for a test φ such that

$$P_0(\varphi = 1) \le \alpha_0$$

and

$$P_f(\varphi = 0) \le \beta_0$$

for all f from the alternative set H_1.

We treat this problem in an asymptotic sense as $\varepsilon \to 0$. The goal is to describe the optimal rate $r(\varepsilon)$ for which testing with the prescribed error probabilities is still possible.

3. Translation into Sequence Space

3.1. Wavelet basis and Model in Sequence Space

Below we pass from the original nonparametric problem to the parametric problem of high dimension, namely to the problem in terms of wavelet coefficients. We follows Donoho and Johnstone (1992).

Assume we are given an orthonormal basis of compactly supported wavelets of $L_2[0,1]$. One may use the construction from Meyer (1991) or Cohen, Daubechies and Vial (1993).

Let $\phi_{j,k}, \psi_{j,k}$ be a system of compactly supported orthogonal wavelets ($\sup \phi \subseteq [-0, A]$ and $\sup \psi \subseteq [-0, A]$). We suppose that ϕ and $\psi \in C^m$, $m = [s]$ (here $[\cdot]$ is an integer part). This implies (cf. Daubechies, 1992, ch. 7) that $\psi(x)$ has at least m vanishing moments.

Let j_0 be such that $2^{j_0} > A + 1$. It has been shown in Cohen, Daubechies and Vial (1993) and Cohen et al. (1993) that an orthogonal wavelet basis on $[0,1]$ can be constructed by retaining $\psi_{j,k}$ and $\phi_{j,k}$ as the interiour wavelets and scaling functions and adding adapted edge wavelets and scaling functions. These edge elements are tailored so that the total number is exactly 2^j at resolution j. For the sake of simplicity we use the same notation for the edge corrected and original functions.

This construction provides an unconditional basis for the $B^s_{p,q}[0,1]$ space for $s > m$, $sp > 1$.

It is useful to use for $\phi_{j_0,k}$ also the notation ψ_k, $k = 1, \ldots, 2^{j_0}$.

Denote by $\mathcal{J}$ the set of resolution levels for the considered wavelet basis,

$$\mathcal{J} = \{j \ge j_0\}$$

and let $\mathcal{I}_j$ be the index set for j th level,

$$\begin{aligned} \mathcal{I}_{j_0} &= \{k: k=1,\dots,2^{j_0}\}\bigcup\{(j_0,k): k=1,\dots,2^{j_0}\}, \\ \mathcal{I}_j &= \{(j,k): k=1,\dots,2^j\}. \end{aligned}$$

By $\mathcal{I}$ we denote the global index set for the considered basis, $\mathcal{I}=\{\mathcal{I}_j,\ j\in\mathcal{J}\}$.

Let now $X_I,\ I\in\mathcal{I}$ be empirical wavelet coefficients for the model (1.1),

$$X_I=\int_0^1\psi_I(t)dX(t).$$

The decomposition (1.1) is reduced to

$$X_I=\theta_I+\varepsilon\int_0^1\psi_I(t)dW(t)$$

Now the original functional model (1.1) is translated into the sequence space model

$$X_I=\theta_I+\varepsilon\xi_I, \qquad I\in\mathcal{I} \tag{3.1}$$

where $\xi_I=\int\psi_I dW$ are standard normal and independent for different I.

This translation is justified by the following two (isometric) properties, cf. Triebel, 1992, p. 240:

(ISO1) For any function $f\in L_2[0,1]$

$$\|f\|^2=\|\theta\|^2:=\sum_{\mathcal{I}}\theta_I^2, \tag{3.2}$$

(ISO2) If $m\ge\sigma$, then there are two constants C_1 and C_2 such that

$$C_1\|f\|_{B^\sigma_{p,q}}\le\|\theta\|_{b^\sigma_{p,q}}\le C_2\|f\|_{B^\sigma_{p,q}},$$

where

$$\|\theta\|_{b^\sigma_{p,q}}=\begin{cases}\left\{\sum\limits_{j\ge j_0}\left[2^{js}\left(\sum\limits_{\mathcal{I}_j}|\theta_I|^p\right)^{1/p}\right]^q\right\}^{1/q}, & q<\infty,\\ \sup\limits_{j\ge j_0}\left\{2^{js}\left(\sum\limits_{\mathcal{I}_j}|\theta_I|^p\right)^{1/p}\right\}, & q=\infty,\end{cases} \tag{3.3}$$

s being $\sigma+\frac{1}{2}-1/p$.

3.2. Hypothesis Testing Problem in Sequence Space

Now we reformulate the hypothesis testing problem for the sequence space model. We wish to test the null hypothesis

$H_0: \qquad \theta_I=0, \qquad I\in\mathcal{I}.$

The alternative is also expressed in term of wavelet coefficients. The condition on the distance between the null and the alternative is of the form

$$\|\theta\| \geq r(\varepsilon)$$

where $\|\theta\|^2 = \sum_{I \in \mathcal{I}} \theta_I^2$.

Next, given $\tau = (\sigma, p, q, M)$, the smoothness condition $\|f\|_{B^\sigma_{p,q}} \leq M$ is transferred into

$$\theta \in \Theta_\tau = \{\|\theta\|_{b^\sigma_{p,q}} \leq M\}.$$

We define the alternative set for the sequence space model as follows.

$H_1: \qquad \|\theta\|_{b^\sigma_{p,q}} \leq M, \qquad \|\theta\| \geq r(\varepsilon)$.

Let now φ be a test i.e. a two-valued function of the observations X_I, $I \in \mathcal{I}$. As usual, the event $\{\varphi = 0\}$ is treated as accepting the null and $\{\varphi = 1\}$ means that the null is rejected. We measure the quality of a test φ by the corresponding error probabilities.

The first kind error probability is defined by

$$\alpha(\varphi) = P_0(\varphi = 1).$$

Here P_0 means the distribution of observations under the null i.e. if $\theta_I = 0$, $I \in \mathcal{I}$, in (3.1).

The behavior of the test on the alternative set is described by the properties of the corresponding power function $\beta_\theta(\varphi)$,

$$\beta_\theta(\varphi) = P_\theta(\varphi = 0).$$

Namely, for the alternative H_1 we consider the maximal value of the power function on the corresponding alternative set,

$$\beta(\varphi; \tau, r(\varepsilon)) = \sup_{\theta \in \Theta_\tau,\, \|\theta\| \geq r(\varepsilon)} P_\theta(\varphi = 0). \tag{3.4}$$

We are interested in tests φ for which

$$\alpha(\varphi) \leq \alpha_0, \qquad \beta(\varphi; \tau, r(\varepsilon)) \leq \beta_0, \tag{3.5}$$

with a prescribed $\alpha_0, \beta_0 \in (0, 1)$.

The problem of minimax hypothesis testing can be defined as follows: to describe the optimal rate for $r(\varepsilon)$ for which testing with the prescribed error probabilities is still possible i.e. the set of tests satisfying (3.5) (at least in an asymptotic sense) is nonempty.

3.3. Minimax rate

Now we present the results which describe the minimax rate of testing for the problem in the sequence space.

Theorem 3.1. *Let, given* $\tau = (\sigma, p, q, M)$ *with* $\sigma p > 1$ *,*

$$r_\tau(\varepsilon) = M^{\frac{1}{4\sigma''+1}} \varepsilon^{\frac{4\sigma''}{4\sigma''+1}} \tag{3.6}$$

where

$$\sigma'' = \sigma - \left(\frac{1}{2p} - \frac{1}{4}\right)_+ = \min\{\sigma, \sigma - 1/(2p) + 1/4\}. \tag{3.7}$$

Then for any $\alpha_0 > 0$ *and* $\beta_0 > 0$ *there exist a constant* $c_1 = c_1(\sigma, p, q, \alpha_0, \beta_0)$ *and a test* φ^* *such that*

$$P_0(\varphi^* = 1) \le \alpha_0 + o_\varepsilon(1) \tag{3.8}$$

and

$$\sup_{\theta\in\Theta_\tau,\, ||\theta||\ge c_1 r_\tau(\varepsilon)} P_\theta(\varphi^* = 0) \le \beta_0 + o_\varepsilon(1).$$

Here $o_\varepsilon(1)$ *means a sequence temding to zero as* $\varepsilon \to 0$*.*

The structure of the test φ^* is explained below in this section. Now we present the lower bound which states rate optimality of this test.

Theorem 3.2. *Let* τ *and* $r_\tau(\varepsilon), \alpha_0, \beta_0$ *be as above. Then there exists a constant* $c_2 = c_2(\sigma, p, q, \alpha_0, \beta_0)$ *such that for any test* φ *satisfying (*3.8*)*

$$\sup_{\theta\in\Theta_\tau,\, ||\theta||\ge c_2 r_\tau(\varepsilon)} P_\theta(\varphi = 0) \ge \beta_0 - o_\varepsilon(1).$$

3.4. Minimax test

First we restrict the considered set of wavelet coefficients $\mathcal{I}$ by some subset $\mathcal{I}_\varepsilon$. This procedure is typical for statistical analysis based on wavelet technique, see e.g. Donoho and Johnstone (1992).

Define the level j_ε as the minimal integer with

$$2^{j_\varepsilon} \ge \varepsilon^{-2}.$$

Set now

$$\begin{aligned} \mathcal{J}_\varepsilon &= \{j \in \mathcal{J} : j \le j_\varepsilon\}, \\ \mathcal{I}_\varepsilon &= \bigcup_{j\in\mathcal{J}_\varepsilon} \mathcal{I}_j. \end{aligned}$$

It is convenient to introduce also the "normalized" observations $Y_I = \varepsilon^{-1} X_I$ i.e. due to (3.1)

$$Y_I = \varepsilon^{-1}\theta_I + \xi_I.$$

Denote for each $j \in \mathcal{J}$

$$S_j = \varepsilon^{-2} \sum_{\mathcal{I}_j} (X_I^2 - \varepsilon^2) = \sum_{\mathcal{I}_j} (Y_I^2 - 1). \tag{3.9}$$

Given $\lambda > 0$, set also

$$S_j(\lambda) = \varepsilon^{-2} \sum_{\mathcal{I}_j} [X_I^2 \mathbf{1}(|X_I| > \varepsilon\lambda) - \varepsilon^2 b(\lambda)] = \sum_{\mathcal{I}_j} [Y_I^2 \mathbf{1}(|Y_I| > \lambda) - b(\lambda)]. \tag{3.10}$$

Here

$$b(\lambda) = E[\xi^2 \mathbf{1}(|\xi| > \lambda)]$$

and ξ means the standard normal variables.

Given $\tau = (\sigma, p, q, M)$, define the level $J \in \mathcal{J}$ by

$$2^{-J} = \left(\frac{\varepsilon}{M}\right)^{\frac{4}{4\sigma''+1}} \tag{3.11}$$

i.e.

$$J = (\sigma'' + 1/4)\log_2(M/\varepsilon).$$

We assume without loss of generality that the right hand-side of this equality is an integer. Otherwise one can take its integer part. Obviously J depends on ε and J tends to infinity as ε tends to zero. In what follows we assume ε to be small enough and $J > j_0$.

Let $\mathcal{J}_+$ and $\mathcal{J}_-$ be the partition of the level set $\mathcal{J}_\varepsilon$ into two parts over and below J

$$\mathcal{J}_+ = \{j : j_0 \le j < J\}, \qquad \mathcal{J}_- = \{j \in \mathcal{J}_\varepsilon : j \ge J\}.$$

Now put for $j \in \mathcal{J}_-$

$$\lambda_j = 4\sqrt{(j - J + 8)\ln 2}, \qquad j \ge J$$

and introduce the test statistics $T(J)$ by

$$T(J) = 2^{-J/2}\left[\sum_{j\in\mathcal{J}_+} S_j + \sum_{j\in\mathcal{J}_-} S_j(\lambda_j)\right]. \tag{3.12}$$

The test φ^* is defined by

$$\varphi^* = \mathbf{1}\left(T(J) > v(J)\Phi^{-1}(\alpha_0)\right) \tag{3.13}$$

where

$$v^2(J) = 2^{-J+1}\left[2^{j_0} + \sum_{j\in\mathcal{J}_+} 2^j + \sum_{j\in\mathcal{J}_-} 2^j d(\lambda_j)\right] \tag{3.14}$$

and

$$d(\lambda) = \frac{1}{2} E\left[\xi^2 \mathbf{1}(|\xi| > \lambda) - b(\lambda)\right]^2.$$

We finish describing the test φ^* with a few remarks.

Remark 3.1. The test φ^* depends on $\tau = (\sigma, p, q, M)$ and ε but this dependence is only through the value J.

Remark 3.2. Easy to check that $v(J)$ converges as $\varepsilon \to 0$ to the value v with

$$v^2 = 2 + \sum_{k=0}^{\infty} 2^k d\left(4\sqrt{(k+8)\ln 2}\right).$$

Hence this universal constant v can be used in place of $v(J)$ for the test φ^*.

Remark 3.3. The choice of the "thresholds" λ_j of the form $C\sqrt{j - j_{\min}}$ was proposed for the estimation problem in Delyon and Juditsky (1995), see also Donoho et al. (1994).

4. Adaptive testing

In the present section we consider the problem of adaptive testing when the parameters $\tau = (\sigma, p, q, M)$ are unknown.

We consider again the model (3.1) in the sequence space. First we state the phenomenon "lack of adaptability" for this problem i.e. we show that adaptive testing with the same rate is impossible.

Then we describe the optimal adaptive rate of testing. For this we use the notion of "adaptive factor".

We start with the definition of the problem of adaptive testing. Let again the alternative set H_1 be described as above but let the parameter τ be unknown. We assume only that τ belongs to some set $\mathcal{T}$. For each $\tau \in \mathcal{T}$ the power of a test φ is determined by the value $\beta(\varphi; \tau, r(\varepsilon))$ due to (3.4) with $r(\varepsilon) = r_\tau(\varepsilon)$ from (3.6).

The results from the preceding section state that for each τ there is $c_1 > 0$ and a test φ_τ such that $\alpha(\varphi_\tau) = P_0(\varphi_\tau = 1) \le \alpha_0$ and $\beta(\varphi_\tau; \tau, c_1 r_\tau(\varepsilon)) \le \beta_0$ (at least in an asymptotic sense). But now for the problem of adaptive testing we search for an universal test φ_ε such that $\alpha(\varphi) \le \alpha_0$ and $\beta(\varphi_\varepsilon; \tau, c r_\tau(\varepsilon)) \le \beta_0$ for some $c > 0$ and all $\tau \in \mathcal{T}$.

We say that a set $\mathcal{T}$ is *nontrivial* if there are such p, q, M and $\sigma_* < \sigma^*$ that

$$\{\sigma, p, q, M\} \in \mathcal{T}, \qquad \forall \sigma \in [\sigma_*, \sigma^*].$$

The first result shows that adaptive testing is impossible (without loss of power) for any nontrivial set $\mathcal{T}$.

Theorem 4.1. *Let $\mathcal{T}$ be nontrivial. Then for any $c > 0$ and any test φ such that*

$$P_0(\varphi = 1) \le \alpha_0 + o_\varepsilon(1) \tag{4.1}$$

one has

$$\sup_{\tau \in \mathcal{T}} \beta(\varphi; \tau, c r_\tau(\varepsilon)) \ge 1 - \alpha_0 - o_\varepsilon(1).$$

The next question is how to define the optimal adaptive rate. One way to do this was proposed for the similar situation for the estimation problem by Lepski (1990), see also Lepski and Spokoiny (1995) where the problem of an optimal adaptive estimation at a point was solved up to an optimal constant.

We use below another approach based on the notion of *adaptive factor*.

Namely, we search for a sequence $t_\varepsilon \to \infty$ such that testing with the rate $r_\tau(\varepsilon t_\varepsilon)$ will be possible adaptively in $\tau \in \mathcal{T}$.

The next results state that for the problem under consideration the minimal adaptive factor is $(\ln \ln \varepsilon^{-2})^{1/4}$.

Theorem 4.2. *Let*

$$t_\varepsilon = (\ln \ln \varepsilon^{-2})^{1/4}. \tag{4.2}$$

Let then a set $\mathcal{T}$ *be of the form*

$$\mathcal{T} = \{\tau = (\sigma, p, q, M) : \sigma \le \sigma^*,\ 1 \le p \le p^*,\ M_* \le M \le M^*,\ \sigma p > 1\}$$

with some prescribed positive $\sigma^*, p^*, M_* \le M^*$ *.*

Then there exist a constant $c_3 = c_3(\sigma^*, p^*, M_*, M^*)$ *and a test* φ_ε *such that*

$$P_0(\varphi_\varepsilon = 1) = o_\varepsilon(1)$$

and

$$\sup_{\tau \in \mathcal{T}} \sup_{\theta \in \Theta_\tau,\, \|\theta\| \ge c_3 r_\tau(\varepsilon t_\varepsilon)} P_\theta(\varphi_\varepsilon = 0) = o_\varepsilon(1).$$

Theorem 4.3. *Let* t_ε *be as above and* $\mathcal{T}$ *is nontrivial. Then there exists a constant* c_4 *such that for any test* φ *satisfying (*4.1*)*

$$\sup_{\tau \in \mathcal{T}} \sup_{\theta \in \Theta_\tau,\, \|\theta\| \ge c_4 r_\tau(\varepsilon t_\varepsilon)} P_\theta(\varphi_\varepsilon = 0) \ge 1 - \alpha_0 - o_\varepsilon(1).$$

Remark 4.1. Here we meet the degenerate behavior of the adaptive test. One can observe from Theorems 4.2 and 4.3 that the adaptive power of the optimal test φ_ε is close to zero or to 1 depending on the constant factor in the distance between the null and the alternative set.

The similar degenerate behavior appeared in the problem of adaptive estimation at a point, see Lepski and Spokoiny (1995).

4.1. Optimal Adaptive Test

Now we describe the structure of the test φ_ε from Theorem 4.2.

The idea of the test is quite clear. Each set $\tau = \{\sigma, p, q, M\}$ determines a procedure with the corresponding value J. Hence one can construct the family of the tests for all feasible values of this parameter (i.e. for all J from $\mathcal{J}$). Then each test can be applied independently and the whole procedure rejects the null hypothesis if at least one test does. The problem here is that each test has a finite error probability of the first kind and this composite procedure has a large value for this error probability. To avoid this problem one has to take the threshold value with an extra growing factor.

More precisely, let $T(J)$ and $v(J)$ be defined by (3.12) and (3.14) respectively. Define the following test

$$\varphi_\varepsilon = \mathbf{1}\left(\sup_{J\in\mathcal{J}^*} T(J)v^{-1}(J) > 2t_\varepsilon^2\right) \tag{4.3}$$

where t_ε is from (4.2). The index set $\mathcal{J}^*$ is determined by the adaptation range $\mathcal{T}$. Roughly speaking, if $J(\tau)$ is the level for the set τ due to (3.11), then $\mathcal{J}^*$ has to contain all $J(\tau)$, $\tau \in \mathcal{T}$. Let $J^* = J(\sigma^*, p^*, M_*)$ by (3.11). The choice

$$\mathcal{J}^* = \{J \in \mathcal{J}_\varepsilon : J \geq J^*\} \tag{4.4}$$

is a proper one.

4.2. Adaptive test. II

Below we present one more adaptive test which is nearly optimal. The reason to consider this test is twofold. First, this new test is of relatively simple structure. Second, this test is a direct analog of the famous wavelet thresholding estimator with the log-threshold, see Donoho et. al. (1994).

The whole test consists of three subtests. The first one operates locally i.e. for each wavelet coefficient.

Let $\mathcal{I}_\varepsilon$ be the index set of cardinality ε^{-2} containing the first ε^2 elements of $\mathcal{I}$. Define the *local test* $\varphi_{\varepsilon,1}$ by

$$\varphi_{loc} = \mathbf{1}\left(\sup_{I\in\mathcal{I}_\varepsilon} |X_I| > 2\varepsilon\sqrt{\ln \varepsilon^{-2}}\right).$$

This "local test" rejects the null if at least one coefficient X_I is too large to be explained by noise fluctuation.

The two remaining tests are global. They can be naturally treated in terms of "gross and detail" structure. We associate the "gross" terms with the wavelet levels below J^* (see (4.4)) and "detail" terms with levels above J^*.

The second test analyses all "gross" terms simultaneously and the last test analyses the "detail" terms within each wavelet level. Namely, set

$$\varphi_{gross} = \mathbf{1}\left(2^{-J^*/2} \sum_{j\in\mathcal{J}, j<J^*} S_j > 2v_0 t_\varepsilon^2\right)$$

where

$$v_0^2 = 2^{-J^*+1}\left[2^{j_0} + \sum_{j\in\mathcal{J}, j<J^*} 2^j\right] \approx 2$$

and

$$\varphi_{detail} = \mathbf{1}\left(\max_{j\in\mathcal{J}^*} 2^{-j/2-1/2} S_j > 2t_\varepsilon^2\right).$$

Finally, the whole test φ_ε' rejects the null if at least one of φ_{loc}, φ_{gross} and φ_{detail} does,

$$\varphi_\varepsilon' = \max\{\varphi_{loc}, \varphi_{gross}, \varphi_{detail}\}.$$

The properties of this test are described by the following result.

Theorem 4.4. *Let*

$$t_\varepsilon' = (\ln \varepsilon^{-2})^{1/2}. \tag{4.5}$$

Let then a set $\mathcal{T}$ be as in Theorem 4.2.

Then there exist a constant $c_9 = c_9(\mathcal{T})$ such that

$$P_0(\varphi_\varepsilon' = 1) = o_\varepsilon(1)$$

and

$$\sup_{\tau\in\mathcal{T}} \sup_{\theta\in\Theta_\tau, \|\theta\|\ge c_9 r_\tau(\varepsilon t_\varepsilon')} P_\theta(\varphi_\varepsilon' = 0) = o_\varepsilon(1).$$

5. Translation into the Function Space

The results of the previous section for the model in the sequence space and the isometric properties (ISO1) and (ISO2) from Section 3 allow to state the results for the original function model (1.1).

Theorem 5.1. *Let*

$$t_\varepsilon = (\ln\ln \varepsilon^{-2})^{1/4}. \tag{5.1}$$

Let then a set $\mathcal{T}$ be of the form

$$\mathcal{T} = \{\tau = (\sigma, p, q, M) : \sigma \le \sigma^*,\ 1 \le p \le p^*,\ M_* \le M \le M^*,\ \sigma p > 1\}$$

with some prescribed positive $\sigma^*, p^*, M_* \le M^*$.

Let the adaptive test φ_ε from the previous section be applied to the system of the empirical wavelet coefficients X_I for an wavelet basis of regularity $m > \sigma^*$. Then there exists a constant $c_7 = c_7(\sigma^*, p^*, M_*, M^*)$ such that

$$P_0(\varphi_\varepsilon = 1) = o_\varepsilon(1)$$

and

$$\sup_{\tau\in\mathcal{T}} \sup_{f\in\mathcal{B}_\tau, \|f\|\geq c_7 r_\tau(\varepsilon t_\varepsilon)} P_f(\varphi_\varepsilon = 0) = o_\varepsilon(1).$$

where $\mathcal{B}_\tau = \{f : \|f\|_{B^\sigma_{p,q}} \leq M\}$ *for* $\tau = (\sigma, p, q, M)$.

Theorem 5.2. *Let* t_ε *be as above and let* $\mathcal{T}$ *be nontrivial. Then there exists a constant* $c_8 > 0$ *such that for any test* φ *satisfying* (4.1)

$$\sup_{\tau\in\mathcal{T}} \sup_{f\in\mathcal{B}_\tau, \|\theta\|\geq c_8 r_\tau(\varepsilon t_\varepsilon)} P_f(\varphi_\varepsilon = 0) \geq 1 - \alpha_0 - o_\varepsilon(1).$$

The similar result can be formulated for the test φ'_ε from the above.

Theorem 5.3. *Let*

$$t'_\varepsilon = (\ln \varepsilon^{-2})^{1/2}. \tag{5.2}$$

Let then a set $\mathcal{T}$ *be as in Theorem* 5.1.

Then there exist a constant $c_5 = c_5(\mathcal{T})$ *such that*

$$P_0(\varphi'_\varepsilon = 1) = o_\varepsilon(1)$$

and

$$\sup_{\tau\in\mathcal{T}} \sup_{f\in\mathcal{B}_\tau, \|f\|\geq c_5 r_\tau(\varepsilon t'_\varepsilon)} P_f(\varphi'_\varepsilon = 0) = o_\varepsilon(1).$$

5.1. Results for other nonparametric statistical models

In the present section we focus ourselves on the "ideal" model "signal + white noise".

Of course, the statistical practice needs in consideration more realistic models such that density or spectral density function model, regression model with heteroskedastic nongaussian errors etc.

We believe that the idea proposed in the present paper are well applicable to these realistic models but the exact theoretical study beyond the scope of the present paper.

We mention only a few papers which can be helpful for these developments. Brown and Low (1990) stated the equivalence in the Le Cam sense of the "white noise" model and Gaussian regression model. Nussbaum (1993) stated the similar result for density model. Neumann and Spokoiny (1995) showed the equivalence in the estimation problem between the regression model with heteroskedastic nongaussian error and the white noise model. Ingster (1984a, 1984b, 1993) treated the hypothesis testing problem for the L_2-ellipsoidal bodies but for the density and spectral density models. Kerkyacharian and Picard (1993) studied the optimal properties of the wavelet shrinkage procedure for the density model.

Härdle and Mammen studied the problem of testing parametric versus nonparametric regression fit for the case of heteroskedastic errors.

6. Proofs

In this section we give the proofs of Theorems 4.2, 4.3 and 4.4. Theorem 3.2 was stated for the Besov function classes in Lepski and Spokoiny (1995). The result of Theorem 3.1 for the proposed test φ^* can be easily deduced from the proof of Theorem 4.2.

6.1. Proof of Theorem 4.2

First we study the behavior of the test φ_ε under H_0 i.e. for $\theta = 0$.

Let the level sets $\mathcal{J}_\varepsilon = \{j : j_0 \le j \le j_\varepsilon\}$ and $\mathcal{J}^* = \{J : J^* \le J \le j_\varepsilon\}$ be introduced in Section 4. In the next lemma we identify S_j from (3.9) with $S_j(\lambda)$ from (3.10) for $\lambda = 0$.

Lemma 6.1. *The following conditions hold true under H_0:*

(i) *For any $\lambda \ge 0$ and each $j \in \mathcal{J}_\varepsilon$*

$$\begin{aligned} E\, S_j(\lambda) &= 0, \\ E\, S_j^2(\lambda) &= 2^j d(\lambda) \end{aligned}$$

where $d(\lambda)$ is from (3.15) and particularly $d(0) = 1$;

(ii) *The random variables $S_j(\lambda_j)$ are independent for different j and any λ_j;*

(iii) *Uniformly in $j \in \mathcal{J}^*$ and $|t| \le 2 \ln \varepsilon^{-2}$*

$$\frac{P_0(2^{-j/2} S_j > t)}{1 - \Phi(t)} \to 1, \qquad \varepsilon \to 0.$$

Proof. The first two statements follow directly from the definition (3.9). The last statement is an easy consequence of the central limit theorem for i.i.d. random variables. Only the fact is important here that $2^j \to \infty$ uniformly in $j \in \mathcal{J}^*$. $\square$

The next technical result describes the behavior of the test statistics $T(J)$ under H_0.

Lemma 6.2. *The following statements are fulfilled uniformly in $J \in \mathcal{J}^*$:*

(i)

$$\begin{aligned} E\, T(J) &= 0, \\ E\, T^2(J) &= v^2(J); \end{aligned}$$

(ii) *Uniformly in $|t| \le 2 \ln \varepsilon^{-2}$*

$$\frac{P_0(v^{-1}(J) T(J) > t)}{1 - \Phi(t)} \to 1, \qquad \varepsilon \to 0.$$

Proof. The first statement of the lemma can be readily checked using (i) and (ii) of Lemma 6.1. The second statement is again an application of the central limit theorem for independent random variables (see Petrov, 1975). □

The last lemma yields the desirable property of the test φ_ε under H_0. In fact, by (ii)

$$\begin{aligned} P_0(\varphi_\varepsilon = 1) &\le \sum_{J \in \mathcal{J}^*} P\left(T(J) > 2v(J)\sqrt{\ln\ln\varepsilon^{-2}}\right) \le \\ &= \sum_{J \in \mathcal{J}^*} \exp\{-\frac{1}{2}4\ln\ln\varepsilon^2\} \\ &= \frac{\#(\mathcal{J}_\varepsilon)}{(\ln\varepsilon^{-2})^2} \le \frac{\ln\varepsilon^{-2}}{(\ln\varepsilon^{-2})^2} \to 0, \qquad \varepsilon \to 0. \end{aligned}$$

Now we turn to studying the power of the test φ_ε. Let us fix some $\tau = (\sigma, p, q, M) \in \mathcal{T}$ and some $\theta \in \Theta_\tau$ i.e. $\|\theta\|_{b^\sigma_{p,q}} \le M$.

Define the level $J = J(\tau)$ by the equality

$$2^{-J} = \left(\frac{\varepsilon t_\varepsilon}{M}\right)^{\frac{4}{4\sigma''+1}} \tag{6.1}$$

where, recall, $t_\varepsilon = (\ln\ln\varepsilon^{-2})^{1/4}$ and $\sigma'' = \sigma - (1/(2p) - 1/4)_+$. We will examine the behavior of the statistic $T(J)$ under P_θ. The goal is to state that for θ from the alternative set one has with a large P_θ-probability $T(J) > 2v(J)t_\varepsilon^2$ that obviously yields the desirable assertion.

For the proof we use the following decomposition

$$T(J) = E_\theta T(J) + T(J) - E_\theta T(J).$$

Denote

$$\gamma = \frac{1}{r_\tau(\varepsilon t_\varepsilon)}\theta.$$

The condition $\|\theta\|^2 \ge c|r_\tau(\varepsilon t_\varepsilon)|^2$ can be rewritten as

$$\|\gamma\|^2 \ge c.$$

We will show that for $\theta \in \Theta_\tau$ one has

$$E_\theta T(J) \ge \left[\frac{1}{2}\|\gamma\|^2 - c_1(\tau)\right] t_\varepsilon^2 \tag{6.2}$$

with some constant $c_1(\tau)$ depending only on τ and uniformly bounded for $\tau \in \mathcal{T}$. Also we state that for ε small enough

$$\mathrm{Var}_\theta T(J) := E_\theta[T(J) - E_\theta T(J)]^2 \le 4 + \|\gamma\|^2. \tag{6.3}$$

Finally we prove that $T(J)$, being centered and normalized, is asymptotically normal under P_θ. Namely, if

$$\zeta(J) = \frac{T(J) - E_\theta T(J)}{\sqrt{\mathrm{Var}_\theta T(J)}}$$

then uniformly in $|t| < \ln \varepsilon^{-2}$

$$\frac{P(-\zeta(J) > t)}{1-\Phi(t)} = 1 - o_\varepsilon(1). \tag{6.4}$$

Before to prove these statements we explain how they imply the assertion of the theorem. Indeed

$$\begin{aligned} P_\theta(\varphi_\varepsilon = 0) &\le P_\theta(T(J) < 2v(J)t_\varepsilon^2) \le \\ &\le P_\theta\left(E_\theta T(J) + \zeta(J)\sqrt{\mathrm{Var}_\theta T(J)} < 2v(J)t_\varepsilon^2\right) \\ &\le P_\theta\left(-\zeta(J) > \frac{E_\theta T(J) - 2v(J)t_\varepsilon^2}{\sqrt{\mathrm{Var}_\theta T(J)}}\right). \end{aligned}$$

To state the result, by (6.4), it suffices to check that

$$\frac{E_\theta T(J) - 2v(J)t_\varepsilon^2}{\sqrt{\mathrm{Var}_\theta T(J)}} \to \infty, \qquad \varepsilon \to 0.$$

But if $\theta \in \Theta_\tau$ is such that

$$\|\gamma\|^2 = \frac{\|\theta\|^2}{r_\tau^2(\varepsilon t_\varepsilon)} \ge 3c_1(\tau) + 6v(J)$$

then by (6.2) and (6.3)

$$\frac{E_\theta T(J) - 2v(J)t_\varepsilon^2}{\sqrt{\mathrm{Var}_\theta T(J)}} \ge \frac{t_\varepsilon^2(\|\gamma\|^2/2 - c_1(\tau) - 2v(J))}{2 + \|\gamma\|} \to \infty, \qquad \varepsilon \to 0.$$

To state (6.2) and (6.3) we use the following consequence of the "smoothness" condition $\|\theta\|_{b_{p,q}^\sigma} \le M$.

Lemma 6.3. *Let $\theta \in \Theta_\tau$ and $p'' = \min\{2, p\}$ and let λ_j be defined by (3.12), $j \in \mathcal{J}$. Then the following conditions hold*

(i)

$$2^{-J/2} \sum_{j \in \mathcal{J}_-} \sum_{\mathcal{I}_j} \varepsilon^{-2}\theta_I^2 \mathbf{1}(|\theta_I| \le \lambda_j \varepsilon) \le c_2(\tau) t_\varepsilon^{p''};$$

(ii)

$$2^{-J/2} \sum_{j \in \mathcal{J}_-} \sum_{\mathcal{I}_j} \mathbf{1}(|\theta_I| \ge \lambda_j \varepsilon) \le c_3(\tau) t_\varepsilon^{p''};$$

where $c_3(\tau) \le 2$.

(iii) *Uniformly in $J \in \mathcal{J}^*$*

$$\varepsilon^{-2} 2^{-J/2}\left[\|\theta\|^2 - \sum_{j \in \mathcal{J}_\varepsilon} \sum_{\mathcal{I}_j} \theta_I^2\right] \le 2M^2 2^{-J/2} \to 0$$

Proof. Consider first the case $p \le 2$. The condition $\theta \in \Theta_\tau$ yields for each $j \in \mathcal{J}$ (see (3.3)

$$\sum_{\mathcal{I}_j} |\theta_I|^p \le 2^{-jsp} M^p, \tag{6.5}$$

s being $\sigma + 1/2 - 1/p$.

Now

$$\begin{aligned} 2^{-J/2} \sum_{j \in \mathcal{J}_-} \sum_{\mathcal{I}_j} \varepsilon^{-2} \theta_I^2 \mathbf{1}(|\theta_I| \le \lambda_j \varepsilon) &\le 2^{-J/2} \sum_{j \in \mathcal{J}_-} \sum_{\mathcal{I}_j} \varepsilon^{-p} |\theta_I|^p \lambda_j^{2-p} \mathbf{1}(|\theta_I| \le \lambda_j \varepsilon) \le \\ &\le 2^{-J/2} \sum_{j \in \mathcal{J}_-} \varepsilon^{-p} \lambda_j^{2-p} \sum_{\mathcal{I}_j} |\theta_I|^p \le \\ &\le \varepsilon^{-p} 2^{-J/2} \sum_{j \in \mathcal{J}_-} \lambda_j^{2-p} 2^{-jsp} \end{aligned}$$

Note that

$$\sum_{j \in \mathcal{J}_-} \lambda_j^{2-p} 2^{-jsp} \le 2^{-Jsp} \sum_{k=0}^{\infty} \left(4\sqrt{k+8}\right)^{2-p} 2^{-ksp} \le c_2(\tau) 2^{-Jsp}$$

where $c_2(\tau)$ is the latest sum and very roughly $c_2(\tau) \le 48$ as $s \ge 1/2$.

Next, using the definition (6.1) of J and the equality $s + 1/(2p) = \sigma + 1/2 - 1/(2p) = \sigma'' + 1/4$, one gets

$$(M/\varepsilon)^p 2^{-J/2} 2^{-Jsp} = (M/\varepsilon)^p \left(\frac{\varepsilon t_\varepsilon}{M}\right)^{\frac{sp+1/2}{\sigma''+1/4}} = t_\varepsilon^p$$

and (i) is proved for $p \le 2$.

The case $p > 2$ is treated in the same way, substituting everywhere 2 in place of p.

To state (ii) we note that for each j by (6.5)

$$\sum_{\mathcal{I}_j} \mathbf{1}(|\theta_I| \ge \lambda_j \varepsilon) \le \sum_{\mathcal{I}_j} (\lambda_j \varepsilon)^{-p} |\theta_I|^p \le (\lambda_j \varepsilon)^{-p} M^p 2^{-jsp}$$

and further we proceed as above and, moreover, one can easily calculate $c_3(\tau) \le 2$.

It remains to check (iii). Let j_ε be the latest resolution level in $\mathcal{J}_\varepsilon$. Using again (6.5) we obtain for any $j > j_\varepsilon$

$$\sum_{\mathcal{I}_j} \theta_I^2 \le \left[\sum_{\mathcal{I}_j} |\theta_I|^p\right]^{2/p} \le M^2 2^{-2js}.$$

Recall that by definition $2^{-j_\varepsilon} \le \varepsilon^2$ and also the condition $\sigma p > 1$ gives $s > 1/2$. Hence

$$\varepsilon^{-2} 2^{-J/2} \left[\|\theta\|^2 - \sum_{j \in \mathcal{J}_\varepsilon} \sum_{\mathcal{I}_j} \theta_I^2\right] \le M^2 2^{-J/2} \varepsilon^{-2} \sum_{j > j_\varepsilon} 2^{-2js} \le 2M^2 2^{-J/2} \to 0. \tag{6.6}$$

This completes the proof of the lemma. $\square$

Now we are ready to state (6.2). One has

$$E_\theta T(J) = 2^{-J/2}\left[\sum_{j\in\mathcal{J}_+}\sum_{\mathcal{I}_j} E_\theta(Y_I^2-1) + \sum_{j\in\mathcal{J}_-}\sum_{\mathcal{I}_j} E_\theta[Y_I^2\mathbf{1}(|Y_I|>\lambda_j) - b(\lambda_j)]\right]$$

where $Y_I = \varepsilon^{-1}\theta_I + \xi_I$.

The random errors ξ_I are standard normal and obviously

$$E_\theta(Y_I^2-1) = \varepsilon^{-2}\theta_I^2.$$

To estimate the second sum in (6.7) we use the following property of the standard normal law.

Lemma 6.4. *For any* $\lambda > 0$ *and each* y

$$B(y,\lambda) := E(y+\xi)^2\mathbf{1}(|y+\xi|>\lambda) - E\,\xi^2\mathbf{1}(|\xi|>\lambda) \ge \frac{1}{2}y^2\mathbf{1}(|y|>\lambda).$$

Proof. We assume without loss of generality that $y\ge 0$. Easy to see that

$$y\,E\,\xi\mathbf{1}(|y+\xi|>\lambda)\ge 0$$

and

$$E\,\xi^2\mathbf{1}(|y+\xi|>\lambda) - E\,\xi^2\mathbf{1}(|\xi|>\lambda)\ge 0.$$

This yields

$$B(y,\lambda)\ge y^2P(|y+\xi|>\lambda)\ge\frac{1}{2}y^2\mathbf{1}(|y|>\lambda).$$

□

By this lemma for each $j\in\mathcal{J}_-$

$$\sum_{\mathcal{I}_j} E_\theta[Y_I^2\mathbf{1}(|Y_I|>\lambda_j) - b(\lambda_j)] \ge \frac{1}{2}\sum_{\mathcal{I}_j}\varepsilon^{-2}\theta_I^2\mathbf{1}(|\theta_I|>\lambda_j\varepsilon) =$$

$$= \frac{\varepsilon^{-2}}{2}\sum_{\mathcal{I}_j}\theta_I^2 - \frac{\varepsilon^{-2}}{2}\sum_{\mathcal{I}_j}\theta_I^2\mathbf{1}(|\theta_I|<\lambda_j\varepsilon)$$

Now applying Lemma 6.3 we obtain

$$\begin{aligned} E\,T(J) &\ge 2^{-J/2}\varepsilon^{-2}\left[\sum_{j\in\mathcal{J}_+}\sum_{\mathcal{I}_j}\theta_I^2 + \frac{1}{2}\sum_{j\in\mathcal{J}_-}\sum_{\mathcal{I}_j}\theta_I^2 - \frac{1}{2}\sum_{j\in\mathcal{J}_-}\sum_{\mathcal{I}_j}\theta_I^2\mathbf{1}(|\theta_I|<\lambda_j\varepsilon)\right]\ge \\ &\ge \frac{1}{2}\left[2^{-J/2}\varepsilon^{-2}\|\theta\|^2 - M^2 2^{-J/2} - c_2(\tau)t_\varepsilon^2\right]. \end{aligned}$$

The definition (6.1) of J gives by (3.6)

$$2^{-J/2}\varepsilon^{-2} = \varepsilon^{-2}\left(\frac{\varepsilon t_\varepsilon}{M}\right)^{\frac{2}{4\sigma''+1}} = (\varepsilon t_\varepsilon)^{\frac{-8\sigma''}{4\sigma''+1}}M^{\frac{2}{4\sigma''+1}}t_\varepsilon^2 = |r_\tau(\varepsilon t_\varepsilon)|^{-2}t_\varepsilon^2$$

that completes the proof of (6.2).

The next step is estimating $\mathrm{Var}_\theta T(J)$.

Since ξ_I and hence Y_I are independent for different I one gets

$$\mathrm{Var}_\theta T(J) = 2^{-J}\left[\sum_{j\in\mathcal{J}_+}\sum_{\mathcal{I}_j}\mathrm{Var}_\theta Y_I^2 + \sum_{j\in\mathcal{J}_-}\sum_{\mathcal{I}_j}\mathrm{Var}_\theta(Y_I^2\mathbf{1}(|Y_I|>\lambda_j))\right].$$

Obviously

$$\begin{aligned}\mathrm{Var}_\theta Y_I^2 &= E\left(|\varepsilon^{-1}\theta_I+\xi_I|^2 - E|\varepsilon^{-1}\theta_I+\xi_I|^2\right)^2 = \\ &= E\left(2\varepsilon^{-1}\theta_I\xi_I+\xi_I^2-1\right)^2 = \\ &= 4\varepsilon^{-2}\theta_I^2+2.\end{aligned}$$

To estimate the value $\mathrm{Var}_\theta(Y_I^2\mathbf{1}(|Y_I|>\lambda_j))$ we use the following technical assertion.

Lemma 6.5. *For each y and any $\lambda\ge 2$*

$$Var\left(|y+\xi|^2\mathbf{1}(|y+\xi|>\lambda)\right)\le 4y^2+2\mathbf{1}(|y|>\lambda/2)+\lambda^4e^{-\lambda^2/8}.$$

Proof. First we note that for any y,λ

$$\mathrm{Var}\left(|y+\xi|^2\mathbf{1}(|y+\xi|>\lambda)\right)\le \mathrm{Var}|y+\xi|^2 = 4y^2+2.$$

Next, one has readily for $\lambda\ge 2$ and $|y|<\lambda/2$

$$\begin{aligned}\mathrm{Var}\left(|y+\xi|^2\mathbf{1}(|y+\xi|>\lambda)\right) &\le E|y+\xi|^4\mathbf{1}(|y+\xi|>\lambda)\\ &\le E|\lambda/2+\xi|^4\mathbf{1}(|\lambda/2+\xi|>\lambda)\le\\ &\le \lambda^4e^{-\lambda^2/8}\end{aligned}$$

and the lemma follows. □

Applying this result, we get

$$\begin{aligned}\mathrm{Var}_\theta T(J) &\le 2^{-J}\Bigg[\sum_{j\in\mathcal{J}_+}\sum_{\mathcal{I}_j}(4\varepsilon^{-2}\theta_I^2+2)+\\ &\quad+\sum_{j\in\mathcal{J}_-}\sum_{\mathcal{I}_j}(4\varepsilon^{-2}\theta_I^2+2\mathbf{1}(|\theta_I|>\lambda_j\varepsilon/2)+\lambda_j^4e^{-\lambda_j^2/8})\Bigg]\le\\ &\le 4\varepsilon^{-2}2^{-J}\|\theta\|^2+2^{-J+j_0}+2^{-J}\sum_{j\in\mathcal{J}_+}2^{j+1}+\\ &\quad+2^{-J+1}\left[\sum_{j\in\mathcal{J}_-}\sum_{\mathcal{I}_j}\mathbf{1}(|\theta_I|>\lambda_j\varepsilon/2)+\sum_{j\in\mathcal{J}_-}2^j\lambda_j^4e^{-\lambda_j^2/8}\right].\end{aligned}\tag{6.7}$$

Obviously

$$2\sum_{j\in\mathcal{J}_+}2^{j-J}\le 2\sum_{k=1}^{J}2^{-k}<2,\tag{6.8}$$

$$\sum_{j\in\mathcal{J}_-}2^{j-J}\lambda_j^4e^{-\lambda_j^2/8}\le\sum_{j=J}^{\infty}2^{j-J}\left(4\sqrt{j-J+8}\right)^4 2^{-2(j-J+8)}\le\sum_{k=0}^{\infty}2^{-k}=2.\tag{6.9}$$

Also, by (ii) of Lemma 6.3

$$2^{-J/2} \sum_{j \in \mathcal{J}_-} \sum_{\mathcal{I}_j} \mathbf{1}(|\theta_I| > \lambda_j \varepsilon/2) \le 4t_\varepsilon^2 \tag{6.10}$$

and similarly to the above

$$\varepsilon^{-2} 2^{-J/2} \|\theta\|^2 = \frac{\|\theta\|^2}{r_\tau^2(\varepsilon t_\varepsilon)} t_\varepsilon^2 = \|\gamma\|^2 t_\varepsilon^2. \tag{6.11}$$

In view of (6.7) – (6.11), we conclude for ε small enough

$$\mathrm{Var}_\theta T(J) \le 4 \cdot 2^{-J/2} \|\gamma\|^2 t_\varepsilon^2 + 4 + 4 \cdot 2^{-J/2} t_\varepsilon^2 \le 4 + \|\gamma\|^2$$

since $2^{-J/2} t_\varepsilon^2 \to 0$ as $\varepsilon \to 0$ uniformly in $J \in \mathcal{J}_\varepsilon$. The assertion (6.3) follows.

It remains to state asymptotic normality of $\zeta(J)$ in the sense of (6.4). To this end, we note that $\zeta(J)$ is a centered and normalized sum of independent random variables having arbitrary number of moments. Moreover, it is not difficult to check that, the third or forth absolute moment of $\zeta(J)$ is bounded uniformly on ε and the desirable asymptotic normality can be stated by application, for instance, the general results by Amosova (1972).

6.2. Proof of Theorem 4.3

To state the lower bound from Theorem 4.3 we apply the Bayes approach developed by Ingster (1993). The idea is as follows.

We pick in a given nontrivial parameter set $\mathcal{T}$ a finite subset $\mathcal{T}_\varepsilon$ with the cardinality $N_\varepsilon = \#(\mathcal{T}_\varepsilon) \asymp \ln \varepsilon^{-2}$.

Then for each $\tau \in \mathcal{T}_\varepsilon$ we construct a prior measure π_τ such that π_τ is concentrated on the corresponding alternative set $\mathcal{F}_\tau = \{\theta : \|\theta\|_{b_{p,q}^\sigma} \le M, \|\theta\| \ge c r_\tau(\varepsilon t_\varepsilon)\}$,

$$\pi_\tau(\mathcal{F}_\tau) = 1. \tag{6.12}$$

The choice of the constant c here will be made precise below.

The whole prior π_ε is taken of the form

$$\pi_\varepsilon = \frac{1}{N_\varepsilon} \sum_{\mathcal{T}_\varepsilon} \pi_\tau .$$

Let P_{π_ε} denote the Bayes measure for the prior π_ε. Obviously for any test φ

$$\sup_{\tau \in \mathcal{T}} \sup_{\theta \in \mathcal{F}_\tau} P_\theta(\varphi = 0) \ge P_{\pi_\varepsilon}(\varphi = 0).$$

We will show that for a special choice of the set $\mathcal{T}_\varepsilon$ and the priors π_τ, $\tau \in \mathcal{T}_\varepsilon$, one has for c small enough

$$Z_{\pi_\varepsilon} := \frac{dP_{\pi_\varepsilon}}{dP_0} \to 1 \tag{6.13}$$

under P_0-probability as $\varepsilon \to 0$.

But this yields for any test φ (see Lehmann, 1959)

$$P_0(\varphi = 1) + P_{\pi_\varepsilon}(\varphi = 0) = 1 + o_\varepsilon(1)$$

and hence the result of the theorem.

Now we present the construction of the set $\mathcal{T}_\varepsilon$ and the priors π_ε satisfying (6.12) and (6.13).

Let $\mathcal{T}$ be a nontrivial parameter set with the corresponding $\sigma_*, \sigma^*, p, q, M$. To be more definite and to simplify calculation we assume that $p \geq 2$ and $M = 1$. The case $p < 2$ can be considered in the similar way but it requires another structure of the priors π_τ and slightly different technique, cf. Ingster (1993).

Recall that in the case of $p \geq 2$ the adaptive rate is defined as

$$r_\tau(\varepsilon t_\varepsilon) = (\varepsilon t_\varepsilon)^{\frac{4\sigma}{4\sigma+1}}.$$

Let, given $\tau = (\sigma, p, q) \in \mathcal{T}$, the level $j(\tau)$ be defined by the equation

$$2^{-j} = (c\varepsilon t_\varepsilon)^{\frac{4}{4\sigma+1}} \tag{6.14}$$

or

$$j(\tau) = \frac{4}{4\sigma+1} \log_2 (c\varepsilon t_\varepsilon)^{-1} \tag{6.15}$$

with some $c \in (0,1)$.

As usual, if this expression is not an integer, we assume its integer part. Since $j(\tau)$ depends on τ only through σ, we will use also the notation $j(\sigma)$.

Denote

$$\begin{aligned} j_* &= j(\sigma_*), \\ j^* &= j(\sigma^*), \\ \mathcal{J}(\mathcal{T}) &= \{j \in \mathcal{J} : j_* \leq j \leq j^*\} \end{aligned}$$

and define for each $j \in \mathcal{J}(\mathcal{T})$ the value $\tau_j = (\sigma_j, p, q)$ by the equality $j = j(\sigma_j)$ or

$$2^{-j} = (c\varepsilon t_\varepsilon)^{\frac{4}{4\sigma_j+1}}. \tag{6.16}$$

The set $\mathcal{T}_\varepsilon$ consists of τ_j, $j \in \mathcal{J}(\mathcal{T})$. Now we define for each j a prior π_j which is concentrated on the level j. Namely, let $\vartheta = (\vartheta_I, I \in \mathcal{I})$ be a random signal (vector) with $\vartheta_I = 0$ for $I \notin \mathcal{I}_j$ and ϑ_I are iid within $\mathcal{I}_j$ with the binomial distribution of the form

$$\pi_j(\vartheta_I = \pm u_\varepsilon) = 1/2$$

where

$$u_\varepsilon = (c\varepsilon t_\varepsilon)^{\frac{4\sigma+2}{4\sigma+1}}. \tag{6.17}$$

First we check the condition (6.12) for these priors. One has obviously

$$\|\vartheta\|^2 = \sum_{\mathcal{I}_j} u_\varepsilon^2 = 2^j u_\varepsilon^2$$

and by (6.14) with $\sigma = \sigma_j$ and by (6.16)

$$2^j u_\varepsilon^2 = (c\varepsilon t_\varepsilon)^{-\frac{4}{4\sigma+1}+\frac{8\sigma}{4\sigma+1}} = (c\varepsilon t_\varepsilon)^{\frac{8\sigma}{4\sigma+1}} = c' r_\tau^2(\varepsilon t_\varepsilon)$$

with $\tau = \tau_j = (\sigma_j, p, q)$ and $c' = c^{\frac{8\sigma}{4\sigma+1}}$.

Next, in the same way

$$\begin{aligned} \|\vartheta\|_{b_{p,q}^\sigma}^p &= 2^{jsp} \sum_{\mathcal{I}_j} u_\varepsilon^p = \\ &= 2^{(\sigma+1/2-1/p)jp} 2^j u_\varepsilon^p = \\ &= (c\varepsilon t_\varepsilon)^{-\frac{4p(\sigma+1/2)}{4\sigma+1}} (c\varepsilon t_\varepsilon)^{\frac{4\sigma+2}{4\sigma+1}p} = 1. \end{aligned}$$

This means that $\pi_j(\vartheta \in \mathcal{F}_{\tau_j}) = 1$ and (6.12) is proved.

At the next step we evaluate the asymptotic expansion of the log-likelihood $\ln(dP_{\pi_j}/dP_0)$ for each $j \in \mathcal{J}(\mathcal{T})$. Denote

$$l_\varepsilon = c^2 t_\varepsilon^2.$$

Lemma 6.6. *The following expansion holds true uniformly in $j \in \mathcal{J}(\mathcal{T})$ under the measure P_0*

$$\ln \frac{dP_{\pi_j}}{dP_0} - l_\varepsilon S_j + l_\varepsilon^2/2 \to 0, \qquad \varepsilon \to 0, \tag{6.18}$$

and

$$\sup_{|t| \le \ln \varepsilon^{-2}} \frac{P_0(S_j > t)}{1 - \Phi(t)} \to 1, \qquad \varepsilon \to 0.$$

Here $S_j = 2^{-j/2} \sum_{\mathcal{I}_j} (\xi_I^2 - 1)$.

Proof. The similar expansion can be found in Ingster (1993) and we give only a sketch of the proof.

One has easily for the model (3.1) and the prior π_j

$$L_j := \ln \frac{dP_{\pi_j}}{dP_0} = \sum_{\mathcal{I}_j} \ln\left(\frac{1}{2}\exp\{\varepsilon^{-1}u_\varepsilon\xi_I - \varepsilon^{-2}u_\varepsilon^2\} + \frac{1}{2}\exp\{-\varepsilon^{-1}u_\varepsilon\xi_I - \varepsilon^{-2}u_\varepsilon^2\}\right).$$

Using Taylor expansion one has readily

$$L_j = \sum_{\mathcal{I}_j} \left[\frac{1}{2}\varepsilon^{-2}u_\varepsilon^2(\xi_I^2 - 1) - \frac{1}{12}\varepsilon^{-4}u_\varepsilon^4\xi_I^4 + O\left(\varepsilon^{-6}u_\varepsilon^6\xi_I^6\right)\right]$$

Notice now that by the definitions (6.17) and (6.14)

$$\varepsilon^{-2}u_\varepsilon^2 = \varepsilon^{-2}(c\varepsilon t_\varepsilon)^{2\frac{4\sigma+2}{4\sigma+1}} = 2^{-j/2}c^2t_\varepsilon^2 = 2^{-j/2}l_\varepsilon.$$

Then, uniformly in $j \in \mathcal{J}(\mathcal{T})$ by the law of large number

$$\varepsilon^{-4}u_\varepsilon^4 \sum_{\mathcal{I}_j}(\xi_I^4 - 3) = l_\varepsilon^2 2^{-j} \sum_{\mathcal{I}_j}(\xi_I^4 - 3) \to 0$$

and

$$\varepsilon^{-6} u_\varepsilon^6 \sum_{\mathcal{I}_j} \xi_I^6 = l_\varepsilon^3 2^{-3j/2} \sum_{\mathcal{I}_j} \xi_I^6 \to 0$$

as $\varepsilon \to 0$ under the measure P_0.

Finally we remark that $\varepsilon^{-2} u_\varepsilon^2 \sum_{\mathcal{I}_j} (\xi_I^2 - 1) = l_\varepsilon S_j$ and the lemma follows. $\square$

Now we state (6.13). The definition of π_ε yields

$$Z_{\pi_\varepsilon} = \frac{1}{N_\varepsilon} \sum_{j \in \mathcal{J}(\mathcal{T})} Z_{\pi_j}$$

where

$$N_\varepsilon = \#(\mathcal{J}(\mathcal{T})) \approx \left(\frac{4}{4\sigma_* + 1} - \frac{4}{4\sigma^* + 1} \right) \log_2 (c \varepsilon t_\varepsilon)^{-1}$$

and for $c < 1$

$$\frac{1}{N_\varepsilon} e^{l_\varepsilon^2} = \frac{1}{N_\varepsilon} e^{c^4 \ln \ln \varepsilon^{-2}} = \frac{1}{N_\varepsilon} \left(\ln \varepsilon^{-2} \right)^{c^4} \to 0, \qquad \varepsilon \to 0.$$

Now the statement (6.13) follows from the next general assertion.

Lemma 6.7. *Let $(\zeta_{in}, i, n \geq 1)$ be a triangle array of independent random variables on a probability space $(\Omega, \mathcal{F}, P)$ such that*

$$\sup_{i \leq n} \sup_{|t| \leq 2\sqrt{\ln n}} \left| \frac{P(\zeta_{in} > t)}{1 - \Phi(t)} - 1 \right| \to 0, \qquad n \to \infty. \tag{6.19}$$

If the sequence l_n be such that

$$\frac{1}{n} e^{l_n^2} \to 0, \qquad n \to \infty,$$

then the following convergence holds under the measure P

$$\frac{1}{n} \sum_{i=1}^{n} \exp\{l_n \zeta_{in} - l_n^2/2\} \to 1.$$

Proof. The statement of the lemma means the law of large number for the random variables

$$Z_{in} = \exp\{l_n \zeta_{in} - l_n^2/2\}.$$

For this it suffices to state (see Petrov, 1975) that

$$E\, Z_{in} \mathbf{1}(|\zeta_{in}| \leq 2 l_n) \to 1, \qquad n \to \infty,$$

and

$$\frac{1}{n^2} \sum_{i=1}^{n} \mathrm{Var}\left(Z_{in} \mathbf{1}(|\zeta_{in}| \leq 2 l_n) \right) \to 0, \qquad n \to \infty.$$

Using the condition (6.19), one may replace ζ_{in} in these statements by a standard normal ζ and Z_{in} by $Z = \exp\{l_n \zeta - l_n^2/2\}$. To complete the proof it remains to note that

$$E \exp\{l_n \zeta - l_n^2/2\}\mathbf{1}(|\zeta| \le 2l_n) \to 1, \qquad n \to \infty,$$

and

$$n^{-1}\mathrm{Var}Z\,\mathbf{1}(|\zeta| \le 2l_n) \le n^{-1}\mathrm{Var}Z \le n^{-1}e^{l_n^2} \to 0, \qquad n \to \infty.$$

□

6.3. Proof of Theorem 4.4

The proof repeats mostly the line of the proof of Theorem 4.2 but we give an independent proof because this statement seems to be of special interest.

Evidently it suffices to state the result (4.3) for each test φ_{loc}, φ_{gross} and φ_{detail} separately.

First we note that under H_0 all the variables $Y_I = \varepsilon^{-1}X_I$ coincide with ξ_I and are standard normal. Hence

$$\begin{aligned} P_0(\varphi_{loc} = 1) &\le \sum_{\mathcal{I}_\varepsilon} P(|\xi_I| > t'_\varepsilon) \le \\ &\le \#(\mathcal{I}_\varepsilon)\, 2\Phi\left(-2\sqrt{\ln \varepsilon^{-2}}\right) \le \\ &\le \varepsilon^{-2}\, e^{-\frac{1}{2}4\ln\varepsilon^{-2}} \to 0, \qquad \varepsilon \to 0. \end{aligned}$$

Next, by (iii) of Lemma 6.1 one has for ε small enough

$$\begin{aligned} P_0(\varphi_{detail} = 1) &\le \sum_{j \in \mathcal{J}^*} P(2^{-j/2}S_j > 2t_\varepsilon^2) \le \\ &\le \#(\mathcal{J}^*)2\Phi(-2\sqrt{\ln\ln\varepsilon^{-2}}) \le \\ &\le \ln\varepsilon^{-2}\, 2e^{-\frac{1}{2}4\ln\ln\varepsilon^{-2}} \to 0, \qquad \varepsilon \to 0. \end{aligned}$$

Finally, similarly to (ii) of Lemma 6.2 one can easily state asymptotic normality of the sum

$$T_0 = 2^{-J^*/2} \sum_{j=j_0}^{J^*-1} S_j$$

with the parameters $\mathcal{N}(0, v_0^2)$ and thus

$$P_0(\varphi_{gross} = 1) = P(T_0 > 2v_0 t_\varepsilon^2) \to 0, \qquad \varepsilon \to 0.$$

This statement completes the proof of the first assertion of the theorem.

Now we evaluate the error probability of the second kind for the adaptive test φ'_ε. In the sequel we assume that a vector θ observed with noise due to (3.1) belongs to the alternative set i.e. $\theta \in \Theta^\sigma_{p,q}(M)$ with some $\tau = (\sigma, p, q, M)$ and $\|\theta\| \ge c_8 r_\tau(\varepsilon t'_\varepsilon)$.

We will show that the probability of the event $\varphi'_\varepsilon = 0$ is asymptotically small in this case.

We proceed as follows. First we study which information about the vector θ can be extracted from the fact that $\varphi_{loc} = \varphi_{gross} = \varphi_{detail} = 0$.

Then we show that this information is enough to state the desirable result.

Lemma 6.8. *Let* $|\theta_I| > 4\varepsilon\sqrt{\ln \varepsilon^{-2}}$ *for some* $I \in \mathcal{I}_\varepsilon$. *Then*

$$P_\theta = (\varphi_{loc} = 0) = o_\varepsilon(1).$$

Proof. Let us find $\theta_I \geq 4\varepsilon t'_\varepsilon$. Now evidently

$$\begin{aligned} P_\theta(\varphi_{loc} = 0) &\leq P(|X_I| \leq 2\varepsilon t'_\varepsilon) \leq \\ &\leq P(|\theta_I + \varepsilon\xi_I| \leq 2\varepsilon t'_\varepsilon) \leq \\ &\leq P(|\xi_I| > 2t'_\varepsilon) \to 0, \qquad \varepsilon \to 0. \end{aligned}$$

□

Lemma 6.9. *Let* $\sum\limits_{\mathcal{I}_j} \theta_I^2 \geq 4 \cdot 2^{j/2}\varepsilon^2 t_\varepsilon^2$ *for some* $j \in \mathcal{J}^*$. *Then*

$$P_\theta(\varphi_{detail} = 0) = o_\varepsilon(1).$$

Proof. Let j be from the lemma condition. One has by definition of φ_{detail}

$$P_\theta(\varphi_{detail} = 0) \leq P_\theta(2^{-j/2} S_j \leq 2t_\varepsilon^2).$$

Next,

$$S_j = E_\theta S_j + S_j - E_\theta S_j$$

and one has easily

$$\begin{aligned} E_\theta S_j &= \sum_{\mathcal{I}_j} [E(\varepsilon^{-1}\theta_I + \xi_I)^2 - 1] = \sum_{\mathcal{I}_j} \varepsilon^{-2}\theta_I^2, \\ \mathrm{Var}_\theta S_j &= \sum_{\mathcal{I}_j} E[(\varepsilon^{-1}\theta_I + \xi_I)^2 - \varepsilon^{-2}\theta_I^2 - 1]^2 = \sum_{\mathcal{I}_j} [4\varepsilon^{-2}\theta_I^2 + 2]. \end{aligned}$$

Under the lemma condition we get for ε small enough

$$\begin{aligned} E_\theta S_j &\geq 4 \cdot 2^{j/2} t_\varepsilon^2, \\ \frac{E_\theta S_j}{\sqrt{\mathrm{Var}_\theta S_j}} &\geq \frac{4 \cdot 2^{j/2} t_\varepsilon^2}{\sqrt{4 \cdot 2^{j/2} t_\varepsilon^2 + 2^{j+1}}} \geq 2t_\varepsilon^2. \end{aligned}$$

Further, it is not hard to state that the normalized differences

$$\zeta_j = \frac{S_j - E_\theta S_j}{\sqrt{\mathrm{Var}_\theta S_j}}$$

are uniformly in $j \in \mathcal{J}^*$ asymptotically normal in the following sense

$$\sup_{|t| \leq 2\sqrt{\ln \varepsilon^{-2}}} \left| \frac{P_\theta(-\zeta_j > t)}{1 - \Phi(t)} - 1 \right| = o_\varepsilon(1).$$

Now

$$\begin{aligned} P_\theta(S_j < 2^{j/2}\, 2t_\varepsilon^2) &\le P_\theta\left(-\zeta_j \sqrt{\mathrm{Var}_\theta S_j} > E_\theta S_j - 2^{j/2}\, 2t_\varepsilon^2\right) \le \\ &\le P_\theta(-\zeta_j > 2t_\varepsilon^2) \le e^{-t_\varepsilon^4} \to 0, \qquad \varepsilon \to 0, \end{aligned}$$

and the lemma follows. □

To formulate the similar result for the test φ_{gross}, we introduce the following notation. Let θ^* be defined by $\theta_I^* = \theta_I$ for $I \in \mathcal{I}_j$ and $j < J^*$ and $\theta_I^* = 0$ for $I \in \mathcal{I}_j$ $j \ge J^*$. With it

$$\|\theta^*\|^2 = \sum_{j=j_0}^{J^*-1} \sum_{\mathcal{I}_j} \theta_I^2.$$

Lemma 6.10. *If* $\|\theta^*\|^2 \ge 4v_0^2 2^{J^*/2}\varepsilon^2 t_\varepsilon^2$, *then*

$$P_\theta(\varphi_{gross} = 0) = o_\varepsilon(1).$$

Proof. The proof is similar to that of in the preceding case and we omit the details. □

Remark 6.1. The assertions of these three lemmas can be treated in the following way: if one observes $\varphi_\varepsilon' = 0$, then "almost surely" (with probability close to 1)

$$|\theta_I| \le 4\varepsilon t_\varepsilon', \qquad \forall I \in \mathcal{I}_\varepsilon. \tag{6.20}$$

$$\sum_{\mathcal{I}_j} \theta_I^2 \le 4 \cdot 2^{j/2}\varepsilon^2 t_\varepsilon^2, \qquad \forall j \in \mathcal{J}^*, \tag{6.21}$$

$$\sum_{j=j_0}^{J^*-1} \sum_{\mathcal{I}_j} \theta_I^2 \le 4 \cdot 2^{J^*/2} v_0^2 \varepsilon^2 t_\varepsilon^2. \tag{6.22}$$

We will show that these inequalities imply $\|\theta\| \le r_\tau(\varepsilon t_\varepsilon')$. This yields the statement of the theorem in view of Lemma 6.8 – 6.10.

We start with the case $p \ge 2$. For this case the rate r_τ was stated to $r_\tau(\varepsilon) = M^{\frac{1}{4\sigma+1}} \varepsilon^{\frac{4\sigma}{4\sigma+1}}$.

We will see that in this case (6.22) and (6.21) imply $\|\theta\| \le r_\tau(\varepsilon t_\varepsilon)$ i.e. the test φ_ε' provides optimal adaptive rate in any adaptive range $\mathcal{T}$ with $p \ge 2$.

In fact, for each $j \in \mathcal{J}$ the "smoothness" condition $\|\theta\|_{b_{p,q}^\sigma} \le M$ yields

$$\left\{ \sum_{\mathcal{I}_j} |\theta_I|^p \right\}^{1/p} \le M 2^{-js}, \tag{6.23}$$

$s = \sigma + 1/2 - 1/p$.

Since $p \ge 2$ one gets also

$$\left\{ \sum_{\mathcal{I}_j} |\theta_I|^2 \right\}^{1/2} \le M 2^{-js}.$$

Combining with (6.21) and (6.22) we obtain for any $J > J^*$

$$\begin{aligned}
\|\theta\|^2 &\le \left(\sum_{j\le J^*} + \sum_{J^*<j<J} + \sum_{j\ge J}\right) \sum_{\mathcal{I}_j} |\theta_I|^2 \le \\
&\le 4v_0^2 2^{J^*/2}\varepsilon^2 t_\varepsilon^2 + \sum_{j<J} 4\varepsilon^2 t_\varepsilon^2 2^{j/2} + \sum_{j\ge J} M^2 2^{-2js} \le \\
&\le 4v_0^2 2^{J^*/2}\varepsilon^2 t_\varepsilon^2 + 8\varepsilon^2 t_\varepsilon^2 2^{J/2} + 2M^2 2^{-2Js}
\end{aligned}$$

Now we pick J to minimize the last expression. This leads to $J = \left(\frac{2\varepsilon t_\varepsilon}{M}\right)^{4/(4\sigma+1)}$ and with such J

$$\|\theta\|^2 \le 4\left(\frac{2\varepsilon t_\varepsilon}{M}\right)^{\frac{8\sigma}{4\sigma+1}}$$

and the assertion follows for $p \ge 2$.

For the case $p < 2$ we proceed in the similar way but using in addition the result (6.20) from the test φ_{loc}. For any $j \in \mathcal{J}$ we get by (6.23)

$$\sum_{\mathcal{I}_j} \theta_I^2 \le \sum_{\mathcal{I}_j} (4\varepsilon t'_\varepsilon)^{2-p} |\theta_I|^p \le (4\varepsilon t'_\varepsilon)^{2-p} M^p 2^{-jsp}$$

and similarly to the above we obtain for any J

$$\|\theta\|^2 \le 4v_0^2 2^{J^*/2}\varepsilon^2 t_\varepsilon^2 + 8\varepsilon^2 t_\varepsilon^2 2^{J/2} + 2(4\varepsilon t'_\varepsilon)^{2-p} M^p 2^{-jsp}.$$

Now the "optimal" choice of J leads to

$$J \approx \frac{1}{s + 1/(2p)} \log_2 \left(\frac{4\varepsilon t'_\varepsilon}{M}\right)^{-1} = \frac{4}{4\sigma'' + 1} \log_2 \left(\frac{4\varepsilon t'_\varepsilon}{M}\right)^{-1}.$$

With it one gets

$$\|\theta\|^2 \le const. \left(\frac{\varepsilon t'_\varepsilon}{M}\right)^{\frac{8\sigma''}{4\sigma''+1}}$$

and the theorem follows.

References

1. Amosova, N.N. (1972). On limit theorem for probabilities of moderate deviations. *Vestnik Leningrad. Univ.* **13** 5 – 14 (in Russian).
2. Brown, L.D. and Low, M.G. (1990). Asymptotic equivalence of nonparametric regression and white noise. *Technical Report*, Cornell University.
3. Cohen, A., Daubechies, I., Jawerth, B. and Vial, P. (1993). MULTIRESOLUTION ANALYSIS, WAVELETS, AND FAST ALGORITHMS ON AN INTERVAL. *Comptes Rendus Acad. Sci. Paris* (A). **316**., 417–421, 1993.
4. Cohen, A., Daubechies, B. and Vial, P. (1993). WAVELETS ON THE INTERVAL AND FAST WAVELET TRANSFORMS. *Applied and Computational Harmonic Analysis* (A). **1**., 54–81, 1993.
5. I. Daubechies (1992) *Ten Lectures on Wavelets.* SIAM, Philadelphia, 1992.
6. Donoho,D.L. and Johnstone,I.M. (1992). Minimax estimation via wavelet shrinkage *Technical Report 402* Dep. of Statistics, Stanford University.
7. Donoho,D.L. and Johnstone,I.M. (1993). Adapting to Unknown Smoothness via wavelet shrinkage. *Technical Report* Dep. of Statistics, Stanford University.
8. Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1994). Wavelet shrinkage: asymptopia? *J. Royal Statist. Soc.*, to appear.
9. Efroimovich,S. and Pinsker,M.S.(1984) Learning algorithm for nonparametric filtering. *Automation and Remote Control* **11** 1434-1440
10. Ermakov, M.S. (1990). Minimax detection of a signal in a white Gaussian noise. *Theory Probab. Appl.* **35** 667 – 679
11. Golubev,G.K.(1990) Quasilinear estimates of signal in L_2. *Prob. of Info. Transmission* **26** 15-20
12. Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Annals of Statistic* **4** 1926 – 1947
13. Ibragimov,I.A. and Khasminskii,R.Z. (1981). *Statistical Estimation: Asymptotic Theory* Springer, New York.
14. Ingster, Yu.I. (1982). Minimax nonparametric detection of signals in white Gaussian noise *Problems Inform. Transmission* **18** 130 – 140.
15. Ingster, Yu.I. (1984a). Asymptotic minimax testing of nonparametric hypothesis on the distribution density of an independent sample. *Zapiski Nauchn. Seminar. LOMI* **136** 74 – 96. (In Russian)
16. Ingster, Yu.I. (1984b). An asymptotic minimax test of nonparametric hypothesis about spectral density. *Theory Probab. Appl.* **29** 846 – 847
17. Ingster, Yu.I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives. I. *Math. Methods of Statist.* **2** (1993) 85 – 114,
   II. *Math. Methods of Statist.* **3** (1993) 171 – 189,
   III. *Math. Methods of Statist.* **4** (1993) 249 – 268.
18. Kerkyacharian,G. and Picard,D.(1993). Density estimation by kernel and wavelet method, optimality in Besov space *Statistics and Probability Letters* **18** 327–336.
19. Lepski,O.V.(1990). One problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35** N.3 459–470.
20. Lepski,O.V.(1992). Asymptotic minimax adaptive estimation. 2. Statistical model without optimal adaptation. Adaptive estimators *Theory Probab. Appl.* **37** N.3, 468–481.
21. Lepski, O.V., Mammen, E. and Spokoiny, V.G. (1994). Ideal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selection. (submitted in *Ann. Statist.*)
22. Lepski, O.V. and Spokoiny, V.G. (1995a). Optimal Pointwise Adaptive Methods in Nonparametric Estimation. (submitted in *Ann. Statist.*)

23. Lepski, O.V. and Spokoiny, V.G. (1995b). Minimax Nonparametric Hypothesis Testing: The Case of an Inhomogeneous Alternative. (unpublished manuscript)
24. Lehmann, E.L. (1959) *Testing Statistical Hypothesis* Wiley, New York.
25. Neumann, M. and Spokoiny V. (1995). On the Efficiency of Wavelet Estimators under Arbitrary Error Distributions. (to be published in *Math. Methods of Statistics*, 1995).
26. Marron,J.S.(1988) Automatic smoothing parameter selection: a survey. *Empir. Econ.* **13**, 187–208.
27. Nussbaum, M. (1993). Asymptotic equivalence of density estimation and white noise. *Technical Report*, Institute of Applied Analysis and Stochastics, Berlin.
28. Petrov, V.V. (1975) *Sums of Independent Random Variables.* Springer, New York.
29. Poljak,B.T. and Tsybakov,A.B. (1990) Asymptotic optimality of C_p-test for the orthogonal series estimation of regression *Treory Probab. Appl.* **35** N.2 293-306
30. Suslina, I.A. (1993) Minimax detection of a signal for l_q-ellipsoids with a removed l_p-ball. *Zapiski Nauchn. Seminar. SPOMI* **207** 127 – 137. (In Russian).
31. Triebel, H. (1992). *Theory of function spaces II.* Birkhäuser, Basel.

Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstr. 39, 10117 Berlin, Germany, and Institute for Information Transmission Problems, Ermolovoy 19, Moscow 101447, Russia